

BUNDESREPUBLIK DEUTSCHLAND

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



REC'D 25 MAR 2004

WIPO PCT

Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen: 103 13 310.0

Anmeldetag: 25. März 2003

Anmelder/Inhaber: Siemens Aktiengesellschaft,
80333 München/DE

Bezeichnung: Verfahren zur sprecherabhängigen
Spracherkennung und Spracher-
kennungssystem

IPC: G 10 L 15/06

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-
sprünglichen Unterlagen dieser Patentanmeldung.

München, den 12. Februar 2004
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Deizeron

Beschreibung

Verfahren zur sprecherabhängigen Spracherkennung und
Spracherkennungssystem

5

Die Erfindung bezieht sich auf ein Verfahren zur
sprecherabhängigen Spracherkennung mit einem
Spracherkennungssystem, bei dem Sprachäußerungen eines
10 Benutzers trainiert werden und den trainierten
Sprachäußerungen Befehle zugeordnet werden, sowie auf ein
Spracherkennungssystem zur Durchführung des Verfahrens.

15

Ein solches Verfahren ist nach dem Stand der Technik
aufgeteilt in einen Spracherkennungsmodus und einen
Trainingsmodus. Bei dem Spracherkennungsmodus werden
Sprachäußerungen des Benutzers erfasst, woraufhin in einer
Datenbank ein der Sprachäußerung zugeordneter Befehl
aufgefunden wird, sofern die Sprachäußerung eine ausreichende
20 Übereinstimmung mit einer Sprachäußerung zeigt, die zu dem
Befehl gehört und zu einem früheren Zeitpunkt aufgenommen und
gespeichert wurde. Im Spracherkennungsmodus ist eine neue
Zuordnung zwischen einer neuen Sprachäußerung und einem
neuen Befehl nicht möglich. Diese Vorgänge finden vielmehr in
dem Trainingsmodus statt, bei dem der Benutzer
Sprachäußerungen von sich gibt und jeder einzelnen
Sprachäußerung nach ihrer Aufnahme einen Befehl zugeordnet.
Die gewonnene Zuordnung wird in der Datenbank abgelegt.
Zugeordnete Befehle können beispielsweise Anwahlvorgänge für
30 Teilnehmer eines Kommunikationsnetzwerkes oder
Sprachsteuerbefehle sein.

35

Das Verfahren nach dem Stand der Technik hat den Nachteil,
dass das Trainieren neuer Befehle insofern umständlich ist,
dass jedes Mal von dem Spracherkennungsmodus zu dem
Trainingsmodus von dem Benutzer aktiv umgeschaltet werden

muss. Dies wirkt sich auch negativ auf die Akzeptanz von Spracherkennungssystemen im Markt aus.

5 Ausgehend hiervon liegt der Erfindung die Aufgabe zugrunde, ein Verfahren zur sprecherabhängigen Spracherkennung sowie ein Spracherkennungssystem dafür anzugeben, bei dem ein Trainieren neuer Befehle zeitsparend durchführbar ist.

10 Diese Aufgabe wird hinsichtlich des Verfahrens der eingangs genannten Art dadurch gelöst, dass bei Nichterkennung einer Sprachäußerung durch das Spracherkennungssystem dem Benutzer angeboten wird, die Sprachäußerung unmittelbar einem neuen Befehl zuzuordnen.

15 Bei der Durchführung des Verfahrens befindet sich ein Spracherkennungssystem immer im Spracherkennungsmodus, wobei jedoch die Option zur Verfügung steht, bei Nichterkennung einer Sprachäußerung unmittelbar eine neue Befehlszuordnung vorzunehmen. Auf diese Weise ist das Trainieren neuer Befehle
20 in die Spracherkennung selbst integriert und kann dann stattfinden, wenn eine Sprachäußerung nicht erkannt worden ist. Befindet sich beispielsweise der Benutzer gerade in der Situation, dass er einen neuen Befehl für ein Spracherkennungssystem trainieren will, reicht es aus, eine bisher nicht benutzte Sprachäußerung zu artikulieren, wonach das Spracherkennungssystem eine Nichterkennung der neuen Sprachäußerung feststellt und dann die Option anbietet, die neue Sprachäußerung einem neuen Befehl zuzuordnen. Nach
30 Vornahme der Zuordnung kann der Befehl unmittelbar ausgeführt werden.

Bei einer bevorzugten Ausführungsform kann bei Nichterkennung der Sprachäußerung durch das Spracherkennungssystem der Benutzer optional entweder die Sprachäußerung wiederholen
35 oder die Sprachäußerung einem neuen Befehl zuordnen. Diese Ausführungsform trägt der Tatsache Rechnung, dass eine Sprachäußerung gerade außerhalb eines Ähnlichkeitsbereiches

zu einer Sprachäußerung liegen kann, der bereits ein gewünschter Befehl zugeordnet ist. In diesem Fall ist keine Zuordnung einer neuen Sprachäußerung zu einem neuen Befehl beabsichtigt. Vielmehr muss diese Sprachäußerung zur Verknüpfung mit dem bereits trainierten Befehl wiederholt werden.

Hinsichtlich eines Anfangszustandes eines Spracherkennungssystems ist bevorzugt für das Verfahren vorgesehen, dass in dem Fall, wenn noch keiner Sprachäußerung ein Befehl zugeordnet ist, das Spracherkennungssystem nach seiner Aktivierung sofort das Trainieren eines neuen Befehls anbietet. Dies passiert automatisch, wenn das Spracherkennungssystem die erste Sprachäußerung naturgemäß nicht erkennt und die Option anbietet, einen neuen Befehl zu trainieren.

Bei einer weiteren Ausführungsform kann vorgesehen sein, dass bei Nichterkennung einer Sprachäußerung für einen bereits trainierten Befehl durch das Spracherkennungssystem der Benutzer den Befehl auswählen und diesen Befehl die Sprachäußerung zuordnen kann. Dies betrifft den Fall, dass in einer Datenbank, welche die Zuordnungen zwischen Sprachäußerungen und zugehörigen trainierten Befehlen enthält, eine „schlechte“ Version der Sprachäußerung vorliegt, so dass eine Spracherkennung häufig fehlschlägt. In diesem Fall ist es möglich, dem bereits trainierten Befehl eine neue Sprachäußerung zuzuordnen.

Bevorzugt wird für eine Erkennung einer Sprachäußerung ein Sprachmuster erzeugt, das der Sprachäußerung zugeordnet ist. Solche Sprachmuster, denen eine Extraktion wesentlicher Sprachmerkmale der Sprachäußerung zugrunde liegt, werden dann auch in der Datenbank verwendet, die in diesem Fall eine Zuordnung zwischen Sprachmustern und trainierten Befehlen enthält. Nach ihrer Aufnahme wird jede Sprachäußerung in ein Sprachmuster umgesetzt, das dann weiter verarbeitet wird,

beispielsweise für die Entscheidung, ob es erkennbar ist oder nicht, d. h. bereits innerhalb eines Ähnlichkeitsbereiches eines Sprachmusters in der Datenbank vorliegt.

- 5 In diesem Zusammenhang wird es als bevorzugt angesehen, vor einer Zuordnung eines Befehls zu einer Sprachäußerung zu prüfen, ob die Sprachäußerung zuvor gespeicherten Sprachäußerungen ähnlich ist. Auf diese Weise wird vermieden, dass bei einer Spracherkennung Verwechslungen unter
10 verschiedenen Befehlen vorkommen, weil die jeweils zugehörigen Sprachäußerungen einander zu ähnlich sind. Dazu kann, beispielsweise unter Verwendung der Extraktionsmerkmale für ein Sprachmuster, ein zulässiger Ähnlichkeitsbereich definiert werden.

15

- Die oben genannte Aufgabe wird hinsichtlich eines Spracherkennungssystems gelöst durch ein Spracherkennungssystem für ein sprecherabhängiges Erkennen von Sprache mit einer Sprachaufnahmeeinrichtung zur Aufnahme
20 einer Sprachäußerung eines Benutzers des Spracherkennungssystems, einer Suchmaschine, die zum Zugriff auf eine Datenbank ausgebildet ist, die eine Zuordnung zwischen Sprachäußerungen und Befehlen enthält, um einen der Sprachäußerung zugeordneten Befehl aufzufinden, einer Umsetzeinrichtung zum Umsetzen des auf Grund der Sprachäußerung aufgefundenen Befehls, wobei das Spracherkennungssystem derart ausgebildet ist, dass bei Nichterkennung der Sprachäußerung durch das Spracherkennungssystem dem Benutzer angeboten wird, die
30 Sprachäußerung unmittelbar einem neuen Befehl zuzuordnen.

- Ein solches Spracherkennungssystem erlaubt die Durchführung des oben beschriebenen Verfahrens und zeichnet sich gegenüber bekannten Spracherkennungssystemen dadurch aus, dass
35 innerhalb eines Spracherkennungsmodus das Trainieren neuer Befehle ermöglicht wird.

Bevorzugt ist die Sprachaufnahmeeinrichtung mit einem Speicher verbunden, in dem die Sprachäußerung zwischengespeichert wird und der mit der Datenbank zum Einlesen der Sprachäußerung in die Datenbank verbunden ist.

5 Dies ist bei bekannten Spracherkennungssystemen nicht der Fall, denn dort wird für einen Trainingsmodus unmittelbar auf die Datenbank zugegriffen, während in einem Spracherkennungsmodus eine Sprachäußerung für die Funktion der Suchmaschine zwar zwischengespeichert wird, der dann
10 benutze Speicher aber nicht zum Einlesen einer Sprachäußerung in die Datenbank ausgebildet/verknüpft ist.

Vorzugsweise ist eine Merkmalsextraktionseinrichtung zum Erzeugen eines Sprachmusters aus der Sprachäußerung zwischen
15 der Sprachaufnahmeeinrichtung und dem Speicher vorgesehen und das Sprachmuster ersetzt die Sprachäußerung.

Weitere Vorteile und Merkmale des Spracherkennungssystems sind bereits oben anhand der Beschreibung des Verfahrens zur
20 sprecherabhängigen Spracherkennung erläutert worden.

Ein Ausführungsbeispiel der Erfindung wird nachfolgend anhand der Zeichnung noch näher erläutert. Die einzige Figur zeigt ein Ablaufdiagramm eines Verfahrens zur sprecherabhängigen Spracherkennung.

Ein Verfahren zur sprecherabhängigen Spracherkennung mittels eines Spracherkennungssystems wird nunmehr anhand der Figur 1 erläutert. Nach einem Start des Spracherkennungssystems, das
30 beispielsweise als Computersystem mit Anzeigeeinrichtung realisiert ist, wird einem Benutzer zunächst eine geeignete Benutzeroberfläche angezeigt, die auch eine Aktivierung für eine Aufnahme einer Sprachäußerung beinhaltet („Push-to-Talk“-Aktivierung). In einem ersten Verfahrensschritt 1 wird
35 eine Sprachäußerung des Benutzers/Sprechers aufgenommen, und zwar mit Hilfe einer geeigneten Sprachaufnahmeeinrichtung. Mittels einer Merkmalsextraktionseinrichtung wird in einem

zweiten Schritt 2 ein Sprachmuster der Sprachäußerung erzeugt, wobei ein Sprachmuster durch eine Kombination extrahierter, charakteristischer Sprachmerkmale definiert ist. Das Sprachmuster wird in einem Speicher
5 zwischengespeichert.

In einem dritten Schritt 3 erfolgt mittels einer Suchmaschine die Abfrage, ob das erzeugte Sprachmuster in einer Datenbank enthalten ist, die Zuordnungen zwischen Sprachmustern und
10 Befehlen enthält. Diese Datenbank wird in einem Trainingsmodus des Spracherkennungssystems mit Inhalten versehen, wobei der Trainingsmodus in den Vorgang einer Spracherkennung integriert ist. Wird das Sprachmuster als in der Datenbank bereits vorhanden erkannt und der zugehörige
15 Befehl aufgefunden, wird in einem vierten Schritt der Befehl ausgeführt, wonach der Betriebsvorgang des Spracherkennungssystems beendet ist. Dabei erfolgt der Ablauf von dem Schritt 1 bis zu dem Schritt 4 im vorliegenden Ausführungsbeispiel automatisch.

20 Wird in dem dritten Schritt 3 das erzeugte Sprachmuster nicht erkannt, erhält der Benutzer über die Benutzeroberfläche des Computersystems die Option, dem nicht erkannten Sprachmuster bzw. der nicht erkannten Sprachäußerung einen neuen Befehl zuzuordnen. Dies geschieht in einem fünften Schritt 5 des Verfahrens. An dieser Stelle wird das Spracherkennungssystem in einen Trainingsmodus umgeschaltet, sofern die Zuordnung eines neuen Befehls gewünscht ist oder automatisch herbeigeführt wird. Alternativ zu dem fünften Schritt 5 kann
30 der Benutzer mit Hilfe der Benutzeroberfläche auch eine neue Sprachäußerungs-Aufnahme auslösen, so dass zu dem ersten Schritt 1 zurückgekehrt wird, um die Sprachäußerung zu wiederholen.

35 Wird die Zuordnung eines neuen Befehls zu dem nicht erkannten Sprachmuster gewählt, erfolgt in einem sechsten Schritt 6 die Aufnahme einer Sprachäußerung, die der nicht erkannten

Sprachäußerung aus dem ersten Schritt entspricht. Im Anschluss daran wird in einem siebten Schritt 7 aus der in dem sechsten Schritt 6 aufgenommenen Sprachäußerung ein Sprachmuster erzeugt, und zwar in derselben Weise, wie in dem
5 oben erläuterten zweiten Schritt 2.

In einem achten Schritt 8 wird eine Ähnlichkeitsprüfung zwischen dem neuen Sprachmuster aus dem siebten Schritt 7 und dem Sprachmuster aus dem zweiten Schritt 2 durchgeführt.

10 Liegt ein gewünschtes Maß an Übereinstimmung zwischen den beiden Sprachmustern nicht vor, beginnt das Verfahren von neuem, bis ein zufriedenstellendes Ergebnis für die Ähnlichkeit der in dem zweiten Schritt 2 und in dem siebten Schritt 7 erzeugten Sprachmuster vorhanden ist. Dabei können
15 der dritte Schritt 3 und der fünfte Schritt 5 übersprungen werden.

In dem achten Schritt 8 kann auch eine Ähnlichkeitsprüfung dahingehend vorgenommen werden, ob das Sprachmuster der neu
20 aufgenommenen Sprachäußerung einen ausreichenden Abstand gegenüber bereits in der Datenbank vorhandenen Sprachmustern einhält. Im negativen Fall kann der Benutzer aufgefordert werden, eine andere Sprachäußerung zur Zuordnung für einen neuen Befehl zu verwenden. Für diese neue Sprachäußerung beginnt das Verfahren von neuem.

Im Anschluss daran wird in einem neunten Schritt 9 dem im zweiten Schritt 2 erzeugten Sprachmuster ein Befehl zugeordnet, und zwar durch geeignete Auswahl des Benutzers
30 mit Hilfe der Benutzeroberfläche des Spracherkennungssystems. Dazu wird das Sprachmuster aus dem Speicher, in dem es in dem zweiten Schritt 2 zwischengespeichert wurde, ausgelesen, mit dem in Schritt 7 erzeugten Sprachmuster geeignet verknüpft, z. B. durch eine Mittelwertbildung einzelner Eigenschaften
35 beider Sprachmuster, und zusammen mit dem neuen Befehl in die Datenbank geschrieben.

In einem letzten Schritt 10 wird der neu zugeordnete Befehl ausgeführt, wonach der Spracherkennungsvorgang mit integriertem Trainingsmodus abgeschlossen ist.

- 5 Es ist hervorzuheben, dass die in dem vierten und letzten Schritt erfolgende Ausführung eines Befehls mit Hilfe einer Umsetzeinrichtung zum Umsetzen des Befehls vonstatten geht. Bei dem Befehl kann es sich beispielsweise um die Anwahl einer Telefonnummer in einem Kommunikationsnetzwerk oder
10 einen Sprachbefehl handeln, mit dem an ein Netzwerk angeschlossene Geräte gesteuert werden.

- Selbstverständlich kann bei einer vereinfachten Ausführungsform des Verfahrens bei der Zuordnung eines
15 Befehls nach dem neunten Schritt 9 auf die Durchführung der vorangehenden Schritte 6 bis 8 verzichtet werden. Auf diese Weise erfolgt die Zuordnung eines Befehls unmittelbar auf die Abfrage aus dem fünften Schritt 5. Auch ist es möglich, bei der Durchführung des Verfahrens auf die unmittelbare
20 Ausführung des neu trainierten Befehls (zehnter Schritt) zu verzichten.

Patentansprüche

1. Verfahren zur sprecherabhängigen Spracherkennung mit einem Spracherkennungssystem, bei dem Sprachäußerungen eines Benutzers trainiert werden und den trainierten Sprachäußerungen Befehle zugeordnet werden, dadurch gekennzeichnet, dass bei Nichterkennung einer Sprachäußerung durch das Spracherkennungssystem dem Benutzer angeboten wird, die Sprachäußerung unmittelbar einem neuen Befehl zuzuordnen.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass bei Nichterkennung der Sprachäußerung durch das Spracherkennungssystem der Benutzer optional entweder die Sprachäußerung wiederholen oder die Sprachäußerung einem neuen Befehl zuordnen kann.
3. Verfahren nach einem der Ansprüche 1 oder 2, dadurch gekennzeichnet, dass in dem Fall, wenn noch keiner Sprachäußerung ein Befehl zugeordnet ist, das Spracherkennungssystem nach seiner Aktivierung das Trainieren eines neuen Befehls anbietet.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, dass bei Nichterkennung einer Sprachäußerung für einen bereits trainierten Befehl durch das Spracherkennungssystem der Benutzer den Befehl auswählen und diesem Befehl die Sprachäußerung zuordnen kann.
5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, dass für eine Erkennung einer Sprachäußerung ein Sprachmuster erzeugt wird, dass der Sprachäußerung zugeordnet ist.

6. Verfahren nach einem der Ansprüche 1 bis 5,
dadurch gekennzeichnet, dass
vor einer Zuordnung eines Befehls zu einer Sprachäußerung
geprüft wird, ob die Sprachäußerung zuvor gespeicherten
5 Sprachäußerungen ähnlich ist.

7. Spracherkennungssystem für ein sprecherabhängiges Erkennen
von Sprache mit
einer Sprachaufnahmeeinrichtung zur Aufnahme einer
10 Sprachäußerung eines Benutzers des Spracherkennungssystems,
einer Suchmaschine, die zum Zugriff auf eine Datenbank
ausgebildet ist, die eine Zuordnung zwischen Sprachäußerungen
und Befehlen enthält, um einen der Sprachäußerung
zugeordneten Befehl aufzufinden,
15 einer Umsetzeinrichtung zum Umsetzen des auf Grund der
Sprachäußerung aufgefundenen Befehls,
dadurch gekennzeichnet, dass
das Spracherkennungssystem derart ausgebildet ist, dass bei
Nichterkenennung der Sprachäußerung durch das
20 Spracherkennungssystem dem Benutzer angeboten wird, die
Sprachäußerung unmittelbar einem neuen Befehl zuzuordnen.

8. Spracherkennungssystem nach Anspruch 7,
dadurch gekennzeichnet, dass
die Sprachaufnahmeeinrichtung mit einem Speicher verbunden
ist, in dem die Sprachäußerung zwischengespeichert wird und
der mit der Datenbank zum Einlesen der Sprachäußerung in die
Datenbank verbunden ist.

9. Spracherkennungssystem nach einem der Ansprüche 7 oder 8,
dadurch gekennzeichnet, dass
eine Merkmalsextraktionseinrichtung zum Erzeugen eines
Sprachmusters aus der Sprachäußerung zwischen der
Sprachaufnahmeeinrichtung und dem Speicher vorgesehen ist und
35 das Sprachmuster die Sprachäußerung ersetzt.

Zusammenfassung

Verfahren zur sprecherabhängigen Spracherkennung und
Spracherkennungssystem dafür

5

Bei einem Verfahren zur sprecherabhängigen Spracherkennung
mit einem Spracherkennungssystem, bei dem Sprachäußerungen
des Benutzers trainiert werden und den trainierten
10 Sprachäußerungen Befehle zugeordnet werden, wird die Aufgabe,
ein Trainieren neuer Befehle zeitsparend durchzuführen,
dadurch gelöst, dass bei Nichterkennung einer Sprachäußerung
durch das Spracherkennungssystem dem Benutzer angeboten wird,
die Sprachäußerung unmittelbar einem neuen Befehl zuzuordnen.

15

Figur

200219854

